Molecular Psychiatry
https://doi.org/10.1038/s41380-019-0409-1

ARTICLE

AUTS2 isoforms control neuronal differentiation

Galya Monderer-Rothkoff[1] · Nitzan Tal[1] · Marina Risman[1] · Odem Shani[1] · Malka Nissim-Rafinia[1] · Laura Malki-Feldman[1] · Vera Medvedeva[2] · Matthias Groszer[2] · Eran Meshorer[1] · Sagiv Shifman[1]

Received: 27 November 2015 / Revised: 15 March 2019 / Accepted: 18 March 2019

Abstract

Mutations in *AUTS2* are associated with autism, intellectual disability, and microcephaly. AUTS2 is expressed in the brain and interacts with polycomb proteins, yet it is still unclear how mutations in *AUTS2* lead to neurodevelopmental phenotypes. Here we report that when neuronal differentiation is initiated, there is a shift in expression from a long isoform to a short AUTS2 isoform. Yeast two-hybrid screen identified the splicing factor SF3B1 as an interactor of both isoforms, whereas the polycomb group proteins, PCGF3 and PCGF5, were found to interact exclusively with the long AUTS2 isoform. Reporter assays showed that the first exons of the long *AUTS2* isoform function as a transcription repressor, but the part that consist of the short isoform acts as a transcriptional activator, both influenced by the cellular context. The expression levels of PCGF3 influenced the ability of the long AUTS2 isoform to activate or repress transcription. Mouse embryonic stem cells (mESCs) with heterozygote mutations in *Auts2* had an increase in cell death during in vitro corticogenesis, which was significantly rescued by overexpressing the human *AUTS2* transcripts. mESCs with a truncated AUTS2 protein (missing exons 12–20) showed premature neuronal differentiation, whereas cells overexpressing *AUTS2*, especially the long transcript, showed increase in expression of pluripotency markers and delayed differentiation. Taken together, our data suggest that the precise expression of AUTS2 isoforms is essential for regulating transcription and the timing of neuronal differentiation.

Introduction

Neurodevelopmental disorders (NDDs), including intellectual disability (ID) and autism spectrum disorder (ASD), have a strong genetic basis, but most of our current knowledge on risk variants is based on the identification of de novo mutations such as copy number variations and single nucleotide variants [1, 2]. Recent studies identified specific genes that are mutated in individuals with NDDs, but their specificity is very low. First, it is estimated that hundreds of genes can cause NDDs when mutated. Recent studies showed the enrichment of those genes in specific molecular processes of synaptic functions and chromatin regulators [3–7]. Although chromatin regulators may contribute to multiple molecular processes, it has been suggested that the convergence is due to their ability to direct neural progenitor cells (NPCs) toward proliferation and differentiation [8–11]. Second, recent evidence show that the same genes and the same biological pathways may be involved in more than one disorder, as there is a significant overlap between genes with de novo mutations across disorders [12–14]. It is still not clear how mutations in different genes can cause the same disorder and, on the other hand, how mutations within the same gene lead to different disorders.

In this study, we investigated a chromatin regulator gene that has been associated with a range of neurodevelopmental symptoms. Mutations in the activator of transcription and developmental regulator (*AUTS2*, MIM# 607270) are associated with ASD and ID [15–18], but also with other disorders [19–22]. Individuals with mutations in *AUTS2* share common features, including ID, developmental delay, epilepsy, microcephaly, short stature, cerebral palsy, and distinct facial dysmorphisms [22]. Some of the phenotypic variation between individuals with *AUTS2* mutations could be attributed to allelic heterogeneity. It was

Supplementary information The online version of this article (https://doi.org/10.1038/s41380-019-0409-1) contains Supplementary Material, which is available to authorized users.

✉ Sagiv Shifman
sagiv.shifman@mail.huji.ac.il

[1] Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

[2] Inserm, UMR-S839, Sorbonne Universités, Pierre et Marie Curie Université Paris 06, Institut du Fer à Moulin, Paris 75005, France

Published online: 05 April 2019

suggested that the severity of the symptoms is correlated with the proximity of the disruption to the 3′-end of *AUTS2* [22].

The gene expression patterns of *AUTS2* support the involvement of the gene in proliferation and differentiation of NPCs. Bedogni et al. [23] found that the mouse *Auts2* is widely expressed in neuronal progenitors and early differentiating and migrating neurons. Two studies performed a knockdown of *auts2* in *Zebrafish* using morpholinos [22, 24]. The morphant fish showed increased apoptosis in the brain [22, 24], which was rescued by injecting the full-length human *AUTS2* mRNA or a short human isoform of *AUTS2* [22].

The molecular function of the AUTS2 protein has been investigated in two levels. One study found that the full-length AUTS2 isoform is expressed also in the cytoplasm, and that the cytoplasmic AUTS2 regulates Rho family GTPases to control actin dynamics [25]. The conclusion was that AUTS2 is involved in neurite outgrowth and branch formation in neurons, and controls neuronal migration via the Rac1 signaling pathway [25]. Another study described the role of AUTS2 in the nucleus as a part of a polycomb-repressive complex 1 (PRC1) [26]. AUTS2 was previously found to bind two polycomb proteins, PCGF3 and PCGF5 [27], but the major focus was on the PCGF5–AUTS2 complex, which includes RING1A/B, RYBP and its homolog YAF2, and casein kinase 2 (CK2) [26]. Surprisingly, this PCGF5–AUTS2 complex was found to activate transcription. The conversion of a repressive PRC1 complex to transcriptional activator was shown to be the result of the recruitment of CK2 and the interaction of AUTS2 with EP300 [26]. A chromatin immunoprecipitation sequencing (ChIP-Seq) study in mouse embryonic forebrain further supported the role of AUTS2 in gene activation [28].

Here we report that two AUTS2 isoforms display different patterns of expression. The long AUTS2 isoform is expressed in mouse embryonic stem cells (mESCs) but is replaced by the short isoform when neuronal differentiation is initiated. The two transcripts are expressed in the developing human brain but the short isoform is more dominant. In a yeast two-hybrid (YTH) screen, we found that both the short and long isoforms interact with a splicing factor, SF3B1, whereas PCGF3 and PCGF5 interact exclusively with the long isoform. Reporter gene assay suggest that the short isoform has the strongest activation ability, whereas the unique part of the long isoform represses transcription specifically in a neuronal cell line. Disruption of *Auts2* causes cell death during neuronal differentiation, especially when both transcripts are affected. A gene trap that causes the deletion of the 3′ of the gene resulted in accelerated neuronal differentiation, whereas overexpression of *AUTS2*, especially the long isoform, resulted in delayed differentiation.

Materials and methods

Cell culture

Mouse Neuro2a (N2A) cells (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 1% L-glutamine, and 0.5% penicillin–streptomycin in a humidified atmosphere containing 5% CO_2 at 37 °C. To induce differentiation, cells were cultured in low-serum medium (DMEM supplemented with 1% heat-inactivated FBS) and treated with 10 μM retinoic acid (RA, Sigma) for 1–7 days. After the RA treatment, morphological changes were monitored as an indicator of differentiation. The mESC line E14-Tg2A and the SIGTR mESC line AD0706, which has a gene trap within intron 12 of the *Auts2* gene, were cultured on 0.1% gelatin-coated plates in ESC medium (15% ESC-grade FBS, 1 mM sodium pyruvate, 0.1 mM non-essential amino acids, 0.1 mM β-mercaptoethanol, 2 mM glutamine, 100 U ml^{-1} penicillin, and 50 mg ml^{-1} streptomycin, and 1000 U ml^{-1} leukemia inhibitory factor (LIF)). mESCs were grown on feeder layers prepared from mitotically inactivated primary mouse embryo fibroblasts or in a feeder-free culture in 2i medium (ESC medium containing, 1 μM PD0325901 (Peprotech) and 3 μM CHIR99021 (Peprotech)). The mESCs were cultured in humidified atmosphere containing 5% CO_2 at 37 °C. Neuronal differentiation to cortical neurons was performed following the protocol described by Gaspard et al. [29]. This protocol of differentiation is based on a chemically defined medium that is morphogen-free but that contains the Sonic hedgehog inhibitor cyclopamine [30].

RNA expression analysis

Total RNA (0.5–1 μg) was isolated from cells using RNeasy mini kits (Qiagen). For the microarray analysis, the integrity of the RNA was confirmed using the Agilent 2100 bioanalyzer and mRNA was profiled using the Mouse Gene ST 1.0 chip (Affymetrix). Data were normalized using the Expression Console software (Affymetrix). Testing for overrepresentation of functional categories was carried out using ToppGene (toppgene.cchmc.org). As the in vitro corticogenesis cultures consist of a mixture of cells at different stages of their development, we used a low stringent criterion to define differentially expressed (DE) genes between wild-type (WT) and *Auts2*^{GT} cells. Potential DE genes were defined as those that had a 1.5-fold change and a $P < 0.05$ (two-sided *t*-test).

For real-time PCR, cDNA was synthesized using the Superscript III First-Strand kit for reverse transcription-PCR (Invitrogen). Real-time PCR was performed using the iTaq Universal SYBR® Green Supermix (BIO-RAD).

Fluorescence was monitored and analyzed in a Bio-Rad C1000 Thermal Cycler with a CFX96 real-time system. Fold change was determined by the $2^{-\Delta\Delta CT}$ method, using *GAPDH* as the reference gene. All experiments were performed in triplicate. *AUTS2* alternative splicing in adult and embryonic brain samples from human (exons 10 and 12) and mouse (exons 11 and 13) was characterized using PCR with primers designed for exon junctions. PCR was performed with HotStarTaq DNA Polymerase (Qiagen). Primers used for the different assays are listed in Table S1.

Western blotting analyses

Whole-cell lysates were prepared in modified RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA pH 8.0, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) containing protease inhibitors (Sigma-Aldrich). Proteins were extracted by incubation in RIPA buffer with agitation for 1 h at 4 °C. The lysates were cleared by centrifugation (20 min, 12,000 r.p.m., 4 °C). Before loading, Laemmli sample buffer ×5 with 2% β-mercaptoethanol was added to the samples following by boiling at 95 °C for 5 min. Protein samples were loaded on to 8% acrylamide gel and wet transfer was used to move the proteins to a polyvinylidene difluoride membrane. Membranes were blocked in 2% skim milk in phosphate-buffered saline (PBS) and 0.3% Tween-20 (PBST) for 1 h. Primary antibody incubation was overnight in 2% skim milk in PBST 4 °C. The primary antibodies that were used are as follows: anti-AUTS2 (Abcam, ab96326) 1:400, anti-PCGF3 1:100, (Santa Cruz, sc-104578), and anti-β-Tubulin (Abcam, ab6160) 1:10,000. After washes with PBST, membranes were incubated in secondary antibodies (Abcam; ab7083, ab102265) 1:10,000 for 1 h at room temperature (RT) following three washes in PBST. Immunoblottings were visualized using SuperSignal solutions (Pierce) following the manufacturer's instructions.

Immunofluorescence analysis

Cells were seeded to semi-confluence on glass coverslips. The cells were fixed with freshly prepared 4% formaldehyde for 30 min at RT and permeabilized with 0.5% Triton X-100 (Sigma) for 5 min at RT. After blocking with 10% FBS for 1–2 h at RT, the cells were incubated with the primary antibody diluted 1:100 in 5% FBS overnight at 4 °C. Then the coverslips were washed three times with PBS and labeled with the fluorescently tagged secondary antibody (Abcam) diluted 1:500 in PBS for 1 h at RT. The coverslips were mounted with Duolink Mounting Medium with 4′,6-diamidino-2-phenylindole (DAPI) and the slides were examined using an Olympus IX81 microscope with appropriate filters. The primary antibodies that were used are as follows: anti-AUTS2 (Abcam, ab96326), anti-Oct4 (Santa Cruz, sc-8628), anti-Nestin (Sigma, SAB4200394), anti-MAP2 (Sigma-Aldrich, M-1406), and anti-Tubulin (Abcam, ab6160). Secondary antibodies used are as follows: Donkey Anti-Rabbit (Alexa Fluor® 488) (Abcam, ab150065), Goat Anti-Rabbit (Alexa Fluor® 594) (Abcam, ab150084), Donkey Anti-Goat IgG H&L (DyLight® 550) (Abcam,ab96936), and Goat Anti-Rat (Alexa Fluor® 594) (Abcam,ab150168).

Apoptosis analysis

Cells were washed in PBS and stained according to the manufacturer's instructions using the Annexin V–FITC Apoptosis Detection Kit (BD Biosciences). Annexin V–FITC (fluorescein isothiocyanate)-positive cells were scored as early apoptotic cells; cells that were both Annexin V–FITC and propidium iodide (PI) positive were scored as late apoptotic and necrotic cells; and Annexin V–FITC and PI-negative cells were scored as viable. Flow analysis was done with a FACScan (Becton Dickinson). Cell Quest Software was used for data acquisition and analysis. All measurements were done under the same instrument setting and 20,000 cells were analyzed per sample. Three repeats were done for each sample.

Trypan blue staining

To evaluate the number of viable cells, an equal volume of 0.4% Trypan blue was added for 5 min to trypsinized cells at indicated times; live cells were counted using a hemocytometer chamber.

Luciferase reporter assay

pCDNA3.1-GAL4-DB (DNA-binding domain)-positive control plasmid was generated by subcloning the HindIII/HindIII fragment from pGBKT7 into pCDNA3.1 expression vector (HindIII digested). GAL4–AUTS2 fusion protein expression plasmids were generated by subcloning PCR products into pCDNA3.1 using the pGBKT7-AUTS2 plasmids as templates. pGBKT7-AUTS2-ex1-9 was generated by deleting the C terminus of *AUTS2* coding sequence for amino acids 522–1235. For the GAL4-DB fusions, PCGF3, EP300, and PCR-amplified human cDNA was cloned in frame with the GAL4 DNA-binding domain using Gibson cloning. For the *PCGF3* overexpression experiment, amplified human *PCGF3* cDNA was digested with EcoRI and XbaI, and cloned into pCDNA3.1. The primers are outlined in Table S1.

HEK293 and N2A cells were cultured in 48-well tissue culture plates. After overnight incubation, cells were transiently transfected with 0.1 μg UAS-SV40-luc (luciferase reporter) and 0.04 μg pRL-TK (*Renilla* internal control) and

0.25 µg pCDNA3.1-GAL4-DB fusion plasmids, using TransIT-2020. Relative luciferase activity was measured after 48 h using the Dual-Luciferase Reporter system (Promega) according to the manufacturer's protocols, from three independent experiments. To test the effect of *PCGF3* overexpression in HEK293 cells, the cells were transiently transfected with 0.1 µg UAS-SV40-luc (luciferase reporter), 0.04 µg pLR-TK (*Renilla* internal control), 0.15 µg pCDNA3.1-GAL4–AUTS2, and 0.15 µg pCDNA3.1-PCGF3 plasmids, using TransIT-2020. To test the effect of *Pcgf3* knockdown, N2A cells were transiently transfected with small interfering RNA (siRNA) against *Pcgf3* (ON-TARGETplus Mouse SMARTpool, Dharmacon) or control non-targeting siRNA. Twenty-four hours later, the cells were co-transfected with siRNA and the plasmids for the reporter assay using Lipofectamine 2000 Reagent (Invitrogen) according to manufacturer's instructions. Cells were lysed and assayed 24 h post transfection.

YTH screens

YTH experiments were performed with the Matchmaker Gold system (Clontech) using the manufacturer's instructions. The human long variant *AUTS2* cDNA was amplified by PCR with the following primers: 5′-GGGAATTCCATATGGATGGCCCGACGCGGGGCCAT-3′, 5′-TGGAGGCCGAATTCTCGGGCCTCGATATCCTTCAGCGT-3′. The pBluescript *AUTS2* (Open Biosystems, Clone ID 30345831) plasmid was used as template. The short variant was constructed using human cDNA amplified by PCR with the following primers: 5′-GGAATTCATATGCCGACGCCAGCACCTCCCATGGTGCGTACCCCAGGC-3′, 5′-TTATCGGGCCTCGATATCCTTCAGCGT-3′. The PCR products were cloned into the pGBKT7DNA-BD vector in frame with the GAL4 DNA-BD at NdeI/EcoRI sites for the long variant or NdeI/SalI for the short variant, thus creating the bait construct. The pGBKT7- AUTS2 plasmids were transformed into Y2H Gold yeast cells, which were allowed to mate with Y187 yeast cells pre-transformed with the normalized Human Fetal Brain Mate & Plate Library (Clontech), cloned into pGADT7-Rec plasmids. Library plasmids from positive clones that grow on synthetically defined medium lacking tryptophan, leucine, adenosine, and histidine, supplemented with X-α-GAL and survived aureobasidin A selection were rescued, isolated, and sequenced. The pGADT7-PCGF5 and pGADT7-EP300 plasmids were constructed using human cDNA amplified by PCR (the primers are provided in Table S1). The PCR products were cloned into the pGBKT7DNA-BD cut with NdeI/XbaI in frame, with the GAL4 DNA-BD using Gibson Assembly technique. Each plasmid was co-transformed into Y2H Gold strain with the pGBKT7-AUTS2 long- or short-variant plasmid and plated on selective media.

Proximity ligation assay

To detect and localize AUTS2–PCGF3 and AUTS2–SF3B1 interactions, proximity ligation assay (PLA) was performed. N2A cells were grown on glass coverslips, cultured in low-serum medium, and treated with 10 µM RA (Sigma) for 3 days. Cells were fixed in 4% paraformaldehyde for 30 min followed by 5 min permeabilization in 0.5% Triton X-100 and with Duolink blocking buffer for 30 min. The cells were incubated for 1 h at 37 °C with primary antibodies Goat Anti-PCGF3 (sc-104578, Santa Cruz Biotechnologies) or Goat anti-SF3B1 (SAP155) antibody (sc-132445, Santa Cruz Biotechnologies) and Rabbit anti-AUTS2 (Abcam ab96326) diluted in 1:100 in 5% FBS in PBS. Duolink PLA was carried out according to manufacturer'ls instructions using kit reagents (Sigma-Aldrich). We used the PLA probe anti-rabbit MINUS and the PLA probe anti-goat PLUS. The samples were incubated for ligation and amplification. After the washing steps, performed according to the supplier's manual, the slides were mounted using Duolink in situ mounting medium with DAPI. The cells were imaged with fluorescent microscopy. Images were analyzed with the Duolink ImageTool (Sigma-Aldrich).

Genome editing with CRISPR/Cas9

Single-guide RNAs (sgRNAs) targeting exon 1 (5′-CGGGTTAGGAACCGGTGCGGCGG-3′) and exon 9 (5′-GCTGATGGAACTCGGTCCGCAGG-3′) of mouse *AUTS2* gene were designed using CRISPR Design Tool (http://crispr.mit.edu/). The sgRNAs were phosphorylated, annealed, and ligated into pSpCas9(BB)-2A-Puro (PX459) as previously described [31]. Plasmid sgRNAs were transfected into mESCs E14 cells using Lipofectamine 3000 according to manufacturer's specifications. Twenty-four hours post transfection, the cells were selected in the presence of puromycin (1.5 µg/ml) for 48 h, expanded for 10 days, and then transferred to 12-well plates for further expansion. Genomic DNA was extracted and PCR was used to amplify the studied region using the following primers: (exon 10: 5′-GGCAAGCTGCATGCTCTACT-3′ and 5′-CTGAAAGAAAAGGCGTCACC-3′, exon 1: 5′-AAAGCTGGTCATGGCAAATC-3′ and 5′-GGTAGGTGAACGGGGAAGAT-3′). PCR fragments were TA cloned and Sanger sequenced. Selected clones were reseeded into 96-well plates to isolate single-cell-derived colonies.

Lentivirus production and transduction

The overexpression and rescue experiments were conducted by transduction of *AUTS2* constructs to mESCs. The lentivirus was produced by co-transfection ph-AUTS2 (expressing *AUTS2* constructs under the EF1α promoter),

pMD2.G (Addgene), and psPAX2 (Addgene) into HEK293T cells using polyethylenimine (PEI) transfection protocol. The lentiviral supernatants were collected 48–72 h after transfection and were passed through a 0.45 µm filter. The mESCs were incubated with the collected virus-containing medium with 8 mg mL^{-1} polybrene between 5 h and overnight. Twenty-four hours post infection, G418 was added at a final concentration of 0.5 mg/ml for 2 weeks.

Results

Alternative AUTS2 isoforms during in vitro corticogenesis and neuronal differentiation

To study the *AUTS2* gene in cells relevant to embryonic development, we used mESCs and differentiated them to cortical progenitors and neurons (Fig. 1SA) [29]. This model of neurogenesis recapitulates key milestones of early cortical and forebrain development [29]. We studied three different stages of differentiation: before differentiation (D0) and after six (D6) or 12 days (D12) of differentiation. Based on Gaspard et al. [29, 30], D6 is within the neural induction period and the beginning of neurogenesis, whereas D12 is in the neurogenesis stage that corresponds approximately to mouse embryonic day 12 (E12). Nuclear localization of the AUTS2 protein was observed in mESCs cells and during neuronal differentiation (Fig. 1a). Total RNA expression of *Auts2* was not significantly increased during differentiation (1.8-fold change, $P = 0.18$) (Fig. 1b), but western blottings showed a clear isoform switch during neuronal differentiation of mESCs (Fig. 1c). We identified in the western blottings two isoforms: one corresponding to the full-length isoform (140 kDa) and another shorter isoform (~ 110–120 kDa) (Fig. 1c). The long isoform was expressed in mESCs before differentiation (D0) and after 12 days of differentiation (D12), but was completely absent after 6 days of differentiation (D6). In contrast, the short isoform appeared in D6 and continued to be present in D12. The short isoform was previously observed in the developing mouse brain [25].

To test whether the shift in AUTS2 isoforms is specific to neuronal differentiation from mESC, we used N2A cells as another cellular model of neuronal differentiation. N2A cells are derived from spontaneous neuroblastoma of mouse and are capable of differentiating into neuronal-like cells upon RA treatment. In N2A cells, AUTS2 was observed in the nuclei and was concentrated in areas of DAPI light staining, consistent with the localization of AUTS2 to euchromatic regions of the nucleus (Figure S1B). At the RNA level, *Auts2* expression was modestly increased during neuronal differentiation (threefold after 6 days with RA treatment, $P = 4.1 \times 10^{-5}$) (Fig. 1d), but at the protein level, the long AUTS2 isoform was expressed before differentiation and was gradually replaced by the short isoform during neuronal differentiation (Fig. 1e).

The full-length mouse *Auts2* transcript includes 20 exons. There are two short transcripts in the mouse UCSC gene track, which could result in the short AUTS2 protein isoform (Fig. 1f, g). Both short transcripts are expected to result in the same protein. However, one of the short transcripts (uc008zut.1, denoted here as Short1) starts in exon 9 of the long isoform. The other short transcript (uc008zuu.1, denoted here as Short2) also includes exons 6, 7, and 8, and another unique exon—all part of the 5′-untranslated region. To identify transcriptional start sites in the *Auts2* gene, we examined the computationally imputed chromatin states that are associated with active promoters (generated by the ChromHMM algorithm [32]) (Fig. 1g). Two promoter regions were noticed, near the first exon of the long transcript and near the first exon of Short2. These promoters were also supported by associations with CpG islands and Pol II ChIP-Seq data (Fig. 1g). We used exon-level microarray analysis of *Auts2* to study the RNA expression of the two isoforms. Consistent with the western blotting results, the RNA expression of the first four exons of the long *Auts2* transcript, which are not shared with the Short2 transcript, were downregulated in D6 relative to D0 and were upregulated in D12 relative to D6 (Fig. 1h). Exon 11 that was an outlier is alternatively spliced (see below).

We next examined the human *AUTS2* transcripts. The locations of the two promoters for the short and long transcripts were found to be conserved also in humans. Based on RNA-Seq and cap analysis of gene expression data, the human short transcript is similar to mouse, including the unique first exon of Short2, which is not annotated in the human reference sequence (Fig. S2A). To estimate the expression of the two transcripts in the human brain, during different developmental stages, we compared the expression of exon 4 of the long transcript and the expression of exon 1 of the short transcript (both are unique to the long or short transcripts) averaged across different brain regions. Based on this analysis, the short transcript is expressed at higher levels in early stages of brain development (between 8 and 24 post-conception weeks), but both transcripts are expressed in similar low levels in the adult brain (Fig. S2B).

In addition to the alternative promoters producing the long and short *Auts2* transcripts, different transcripts may be generated by alternative exon skipping. In the current genome annotations, two skipped exon events are documented: alternative splicing of exon 12 in human (equivalent to exon 13 in mouse) and exon 11 in mouse (equivalent to exon 10 in human). We used PCR to amplify the alternative exons. We found that alternative skipping of the two

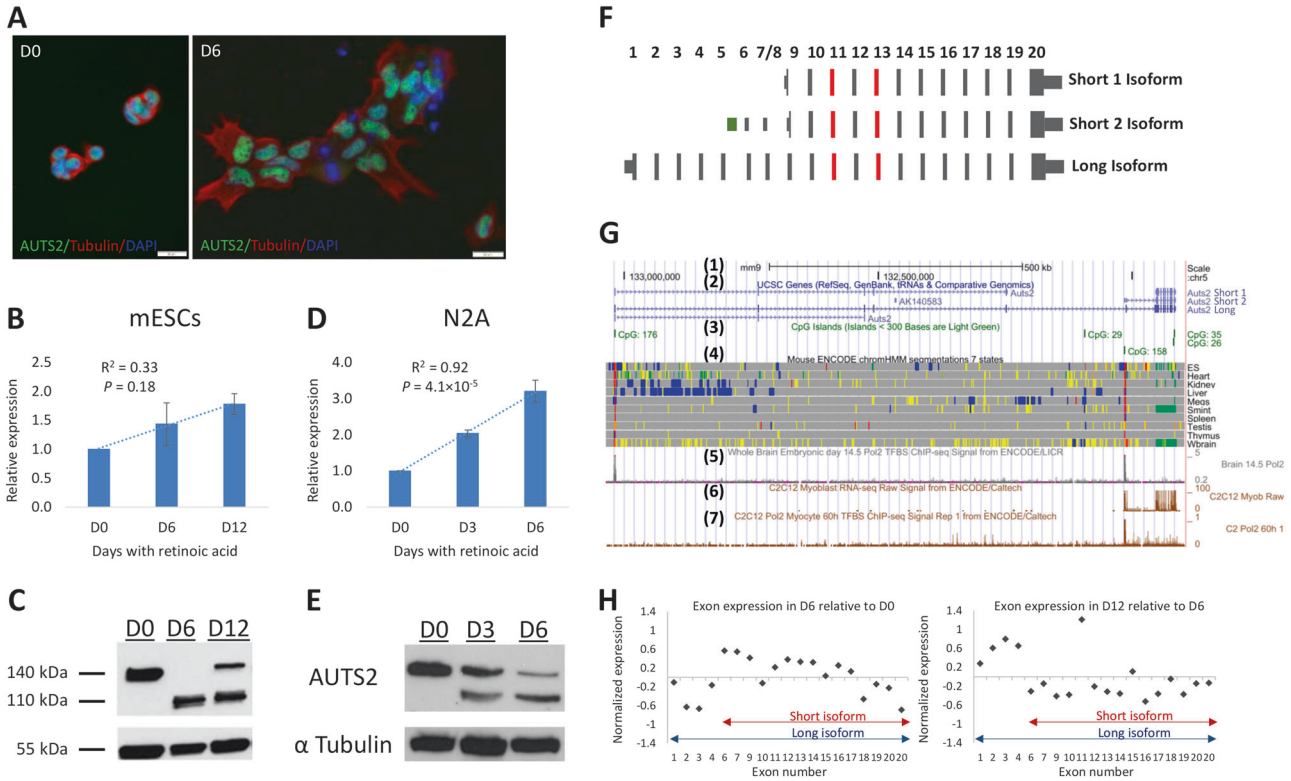

Fig. 1 AUTS2 expression during neuronal differentiation. **a** Immunostaining of AUTS2 in mESCs before and after 6 days in culture shows nuclear localization of the AUTS2 protein. AUTS2, green; Tubulin, red; DAPI, blue. **b** Relative expression of *Auts2* during in vitro corticogenesis of mESCs. The level of *Auts2* RNA expression was measured using quantitative PCR before differentiation (D0) and after six (D6) or 12 days (D6) of differentiation. The primers were designed to detect the two *Auts2* transcripts. Values are the average of three independent measurements with error bars representing the SEM. A nonsignificant linear increase in *Auts2* expression is observed (Pearson's correlation test). R^2 is the proportion of the expression variance explained by the differentiation day. **c** Western blotting analysis of the AUTS2 protein during in vitro corticogenesis. α-Tubulin was used as a control for protein loading. Two isoforms are detected during neuronal differentiation. **d** Relative expression of *Auts2* during retinoic acid-induced neuronal differentiation of N2A cells. A significant linear increase in *Auts2* expression is observed. **e** Western blotting analysis of the AUTS2 protein in undifferentiated N2A cells and after 3 (D3) or 6 days (D6) of retinoic acid treatment. **f** Three transcripts of mouse *Auts2* gene (not to scale). The three transcripts that share the 3′-end of the gene are indicated as Short1, Short2, and Long isoforms. Exons in red (11 and 13) are alternatively spliced. The unique exon of Short2 is in green. **g** The structure of the *Auts2* transcripts and location of predicted promoters obtained from UCSC Genome browser, displaying: (1) position in the genome (mm9); (2) UCSC genes; (3) CpG islands; (4) ChromHMM genomic segmentation with the location of seven states across tissues (Red = H3K4m3 (active promoter), Orange = H3K4m1/3, Yellow = H3K4m1, Light green = H3K4m1 + H3K36m3, Dark green = H3K36m3, Blue = H3K27m3, Gray = Unmarked. Tissues: ES (bruce4), Heart, Kidney, Liver, Megakaryocytes, Small intestine, Spleen, Testis, Thymus, and E14 Whole brain); (5) ChIP-Seq data for RNA Polymerase II (Pol2) in the mouse brain (E14.5) occupying the two predicted promoters for the long and Short2 transcripts; (6) RNA-Seq; and (7) ChIP-Seq data for Pol2 in a myoblast cell line (C2C12), showing expression of the Short2 transcript and Pol2 occupying the promoter region of the Short2 transcript. **h** The expression of the two isoforms during different stages of differentiation was compared at the RNA level based on exon-level analysis of expression microarrays. Values are standard score for the expression differences between D6 and D0 (left plot), and between D12 and D6 (right plot). Exons 1–5 are unique to the long isoform (exon 5 is not measured on the array) and exons 6–20 are shared between the long and Short2 isoforms. In both comparisons, the normalized expression of exon 1–4 is significantly different than the normalized expression of exons 6–20 ($P < 0.05$, two-sided *t*-test)

equivalent exons (10 and 12 in human, 11 and 13 in mouse) occurred in RNA samples from both mouse and human fetal brains, producing four different isoforms (10/12; 10/-; -/12; -/- in human and 11/13; 11/-; -/13; -/- in mouse). In contrast, only two of the variants were amplified in adult human and mouse brains (10/12; 10/- in human and 11/13; 11/- in mouse) (Fig. S3), suggesting that transcripts without exon 10 in human (or 11 in mouse) are specifically expressed in the developing brain but are absent in adult brain.

AUTS2 interacts with PCGF3 and SF3B1

In order to identify proteins that physically interact with the AUTS2 isoforms, we conducted a YTH screen. We used the different AUTS2 isoforms for the screen: (1) the full-length, long isoform (NM_001127231) and (2) a short isoform that consists of exons 9–19 (NM_015570). The YTH screen was done using a GAL4 DNA-binding domain fused to the *AUTS2* gene (as bait) and fetal brain cDNA

library fused to GAL4 DNA-activating domain (as prey). We screened ~106 million diploid yeast clones for the long isoform and 91 million clones for the short isoform. Forty-one positive colonies were identified for the long isoform and 170 for the short isoform. The interactions were confirmed by plasmid rescue from the yeast cells, renewed transformation to yeast cells containing the *AUTS2* expression plasmid. In addition, auto-activation tests were run for all positive plasmids. The plasmid DNA from the YTH interactors was isolated and sequenced. In total, we identified 19 different genes for the long isoform and 27 genes for the short isoform, some represented by multiple clones (Table S2). For each isoform, only five genes were in frame with the prey activation domain, making them much stronger candidates (Table 1).

To identify genes from the list likely to be true interactors of AUTS2 during brain development, we searched for genes that are co-expressed with *AUTS2* in the human developing brain (data from the Allen Brain Atlas). Previous studies showed that co-expression could be used for prediction of protein interactions, with strongly co-expressed genes more likely to have long-lived interactions [33–36]. A weak correlation between *AUTS2* and genes identified in the YTH screen does not necessarily mean that the interaction is false, but it may be less relevant for the developing brain. For example, *PCGF5*, which was previously found to interact with AUTS2 [26], shows a negative correlation with *AUTS2* in the developing brain ($r = -0.78$), suggesting that AUTS2-PCFG5 complex may be irrelevant for brain development, but is likely functional in other tissues and cells. To examine the possibility that genes co-expressed with AUTS2 during different stages of brain development are functionally related, we performed a gene-set enrichment analysis for genes co-expressed with *AUTS2* (using the ToppGene Suite for biological process and human phenotypes [37]) (Table S3). The list included 1762 genes co-expressed with AUTS2 (expression correlation of $r^2 > 0.5$). Notably, the significant enrichment included human phenotypes that are observed in individuals with *AUTS2* mutations and highly relevant terms. The significant phenotypes included: Aplasia/Hypoplasia involving the central nervous system, Global developmental delay, Neurodevelopmental delay, Abnormality of the skull, and Microcephaly (Bonferroni corrected $P < 1 \times 10^{-7}$). The biological processes included chromatin modification, RNA processing, RNA splicing, and negative regulation of gene expression (Bonferroni corrected $P < 1 \times 10^{-10}$). Among the genes identified in the YTH screen, only five were in correlation with AUTS2 ($r^2 > 0.5$) but only three of them are in frame ($r^2 > 0.5$, Table S2). As AUTS2 is mainly a nuclear protein [23] and the short isoform is exclusively found in the nucleus [25], we focused on the two genes that encode nuclear proteins: (1) polycomb group ring finger 3 (*PCGF3*) and (2) splicing factor 3b subunit 1 (*SF3B1*). *PCGF3* is a component of the polycomb group (PcG) multiprotein complex, which was previously found to interact with AUTS2 [26]. *SF3B1* is a splicing factor that belongs to the U2 small nuclear ribonucleoproteins complex. Previous studies showed that SF3B1 interacts with PcG proteins and is essential for the repression of polycomb-mediated repression of *Hox* genes [38], as well as interacts directly with chromatin, mediating chromatin-splicing interactions [39].

Table 1 Genes interacting with AUTS2 based on YTH screen

Genes	No. of clones	AUTS2 isoform	Co-expressed with AUTS2*
PCGF3	1	Long	Yes
PRDX3	2	Long	No
MFSD10	1	Long	No
PGS1	1	Long	No
LRPPRC	1	Long	No
F2R	1	Short	Yes
SF3B1	1	Short	Yes
CHAC1	2	Short	No
ACAA1	2	Short	No
C11orf1	1	Short	No

*Expression correlation of $r^2 > 0.5$ in developing human brain

To validate the interactions and their physiological relevance, we used Duolink in situ PLA with N2A cells. PLA was used to validate the AUTS2–PCGF3 and AUTS2–SF3B1 interactions (Fig. 2). The PLA assay enables the detection of protein interactions as an individual fluorescent dot. PLA was performed in N2A cells after 3 days of differentiation, when both isoforms are expressed. In both experiments, we observed numerous fluorescent spots concentrated within cell nuclei (Fig. 2a). The number of fluorescent dots per nucleus was significantly higher when using both antibodies relative to the control experiment, when one of the antibodies was omitted (Fig. 2b, c). Thus, the PLA experiments support the interaction of AUTS2 with PCGF3 and SF3B1.

The AUTS2 antibody used for the PLA experiments does not distinguish between the long and short isoforms. As we identified the interaction of PCGF3 in the screen for the long isoform and SF3B1 in the screen for the short isoform, we designed YTH experiment to test the interaction between the two AUTS2 isoforms and PCGF3 or SF3B1. As AUTS2 was reported to interact with another polycomb protein, PCGF5 [27], we included this gene also in the experiment. Based on the YTH experiment, the short isoform does not interact with the polycomb proteins, PCGF3 and PCGF5, whereas SF3B1 interacts with both the long and short isoforms (Fig. 2d).

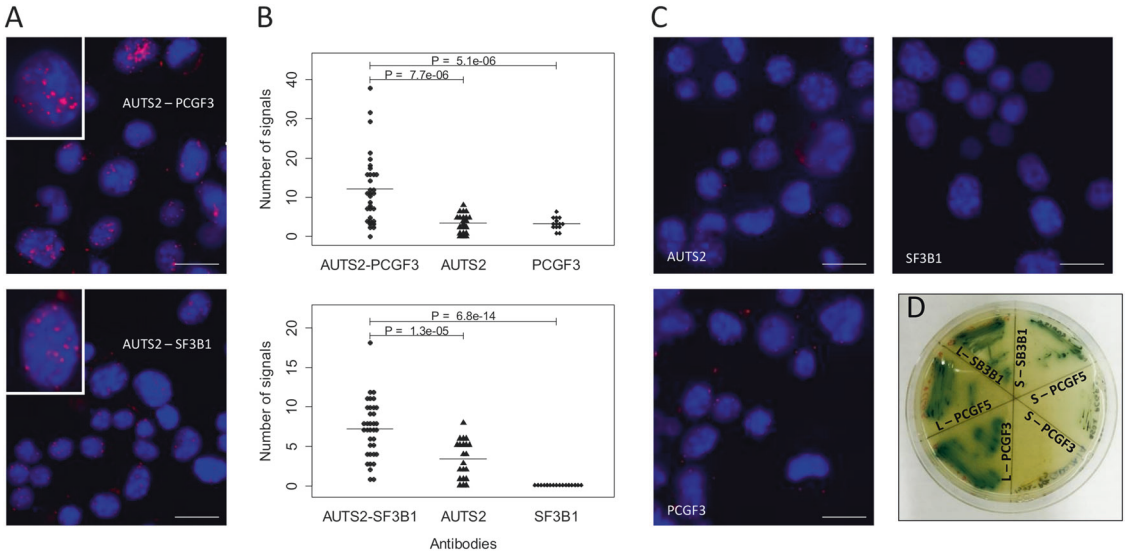

Fig. 2 Validation of protein-protein interactions using the Duolink in situ proximity ligation assay (PLA). **a** Fluorescence microscopy images showing protein interactions (red signal) of AUTS2 with PCGF3 (top) and with SF3B1 (bottom). The inset is an enlargement of nuclei with multiple interaction signals. **b** Quantification of the number of dots per nuclei in the experiments using two primary antibodies relative to the control when one primary antibody was omitted. Each point in the bee swarm plot is a value from one cell. The mean is presented by a horizontal line. Significance was calculated by a two-tailed *t*-test. **c** Fluorescence microscopy images of the controls. **d** α-Galactosidase activity (blue) in streaks of yeast cells on a culture plate as readouts of protein–protein interactions using the yeast two-hybrid system. The yeast two-hybrid system was used to test the interaction between the long (L) and short (S) AUTS2 isoforms and PCGF3, PCGF5, and SF3B1

AUTS2 isoforms are involved in transcriptional activation and repression

A recent work by Gao et al. [26], which studied the AUTS2-PCGF5 complex, showed that AUTS2 is involved in gene activation. Using HEK293 cells, they showed that both AUTS2 and PCGF5 could activate transcription in a reporter assay. Transcriptional activation by AUTS2 was shown to be dependent on PCGF5 [26]. As discussed above, the expression of AUTS2 is negatively correlated with PCGF5 during embryonic stages of brain development, whereas it is positively correlated with PCGF3. To examine the possibility that AUTS2 is involved in gene repression or activation, we used both the human long and short isoforms of AUTS2 fused to the GAL4 DNA-binding domain, as well as PCGF3, in a luciferase reporter assay with HEK293 and undifferentiated N2A cells (Fig. 3a). Western blottings showed that the endogenous AUTS2 and PCGF3 proteins are expressed in N2A but not in HEK293 cells (Fig. 3b). We tested GAL4-P300 construct as a positive control and found a similar transcription activation in both HEK293 and N2A cells (8.1- and 8.8-fold change; $P = 0.00072$, 1.2×10^{-5}, respectively). Consistent with previous results [26], luciferase activity was significantly increased in HEK293 GAL4–AUTS2$_{Long}$ cells (12-fold, $P = 0.00053$), as well as in HEK293 GAL4–PCGF3 cells (3.5-fold, $P = 0.0057$). Notably, the short isoform (GAL4–AUTS2$_{Short}$) showed a strong activation (36-fold, $P = 3.2 \times 10^{-5}$), which was significantly higher than the long AUTS2 isoform ($P = 0.00027$) or PCGF3 ($P = 4.8 \times 10^{-5}$). A construct that included the part that is specific to the long *AUTS2* transcript (human exons 1–8 and part of exon 9, GAL4–AUTS2$_{ex1–9}$) did not show a significant effect ($P = 0.68$) (Fig. 3c).

Surprisingly, a mirror image was found in N2A cells. The most significant effect was a decrease in luciferase activity for the GAL4–AUTS2$_{ex1–9}$ construct (15.7-fold, $P = 1.9 \times 10^{-6}$), followed by a much lower repression for the long isoform (2.0-fold, $P = 0.00081$, GAL4–AUTS2$_{Long}$) and the short isoform (1.5-fold, $P = 0.028$, GAL4–AUTS2$_{short}$) (Fig. 3d). This suggests that the first exons that are specific to the long AUTS2 isoform may function as a transcription repressor, but the part that consists of the short isoform can act as a transcriptional activator, depending on the cellular context.

To test the possibility that the transcription activation or repression by AUTS2 depends on PCGF3, we overexpressed PCGF3 in HEK293 cells (where it is not expressed endogenously) (Figure S4A) and used siRNAs to knock down PCGF3 in N2A cells (Figure S4B). We then measured the ability of the long AUTS2 isoform to activate or repress transcription in these cells. In HEK293 GAL4–AUTS2$_{Long}$ cells, the overexpression of PCGF3 resulted in a significant shift from transcription activation in the untreated cells (9.4-fold increase in luciferase activity) to repression in cells overexpressing PCGF3 (4.3-fold

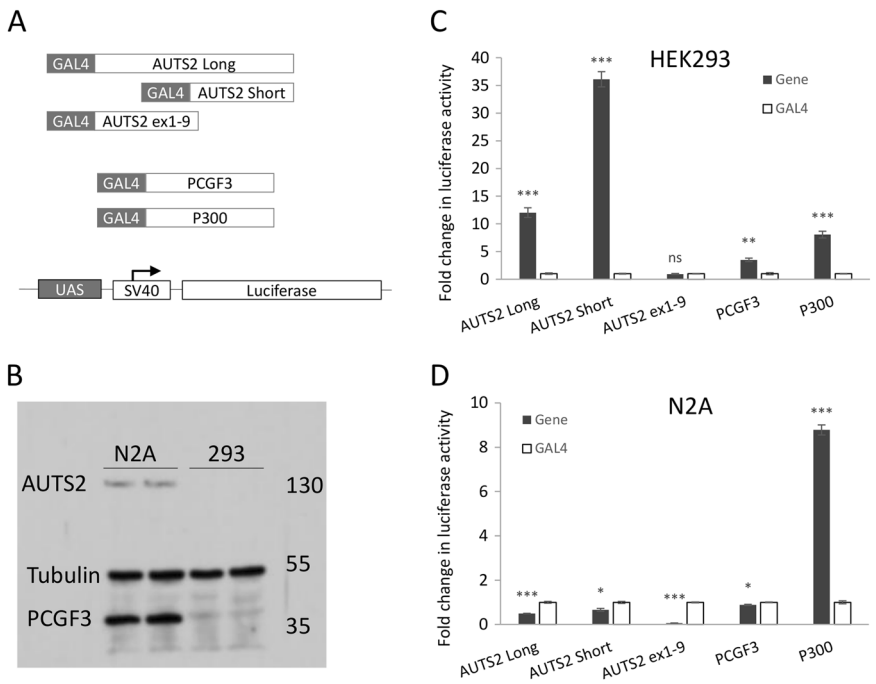

Fig. 3 Dual function of AUTS2 isoforms as both activator and repressor of transcription. **a** Schematic representation (not to scale) of the GAL4 fusion proteins and a reporter gene containing GAL4-binding sites (UAS) located upstream of the SV40 promoter and luciferase gene (UAS-SV40-luc). **b** Western blotting with antibodies against AUTS2, PCGF3, and α Tubulin as a loading control shows that AUTS2 and PCGF3 expressed in N2A cells but not in HEK293. **c**, **d** Luciferase activity was measured in HEK293 and N2A cells expressing GAL4, and GAL4 fused to the long AUTS2 isoform ($AUTS2_{Long}$), the short AUTS2 isoform ($AUTS2_{Short}$), the N terminus of AUTS2 ($AUTS2_{ex1-9}$), PCGF3, and P300 as a positive control. Values are the average of three measurements (±SEM) of the fold change increase in luciferase activity relative to average activity in GAL4 control (ns, not significant; $*P < 0.05$; $**P < 0.01$, $***P < 0.001$; two-sided *t*-test)

decrease in luciferase activity) ($P = 0.0020$) (Figure S4C). In N2A GAL4–$AUTS2_{Long}$ cells, the knockdown of PCGF3 by siRNA resulted in loss of transcription repression (1.25-fold increase in luciferase activity, $P = 0.27$), which was significantly different from untreated cells or cells transfected with control siRNA (~3.0-fold decrease in luciferase activity, $P = 0.00077$ and $P = 0.00036$, respectively) (Figure S4D).

Mutations in Auts2 cause increased cell death during in vitro corticogenesis

To study the molecular and cellular effects of *Auts2* haploinsufficiency, we generated mutations in *Auts2* using CRISPR/Cas9. We used one sgRNA targeting the first exon in attempt to mutate only the long isoform and another sgRNA targeting exon 10, which is included in both the long and the short isoforms. We isolated two clones with heterozygote mutations, one with a deletion of two bases in exon 1 ($Ex1^{+/-}$) and another clone with a deletion of 38 bp in exon 10 ($Ex10^{+/-}$), both predicted to cause a frame shift and premature stop codons (Figure S5). The mutations were confirmed by sequencing (Figure S5A-B), western blotting (Fig. 4a), and immunofluorescence staining (Fig. 4b, c). In addition, we studied mESCs carrying a gene trap ($Auts2^{GT}$) located in intron 11, which results in a truncated *Auts2* transcript missing exons 12–20 (the region responsible for transcription activation) (Figure S5C). As truncated AUTS2 protein is not recognized by the AUTS2 antibody, we used exon-level analysis of expression (based on microarray data) to confirm the gene-trap effect. The analysis showed a higher level of expression in the founder mESC line than in the gene-trap line downstream to the gene trap (Fig. S5C). This reduced expression level was similar to the pattern of expression that we previously observed in an individual that carried a duplication disrupting the *AUTS2* gene [18]. Western blotting and immunofluorescence staining with the AUTS2 antibody showed a reduction in expression of AUTS2 in $Auts2^{GT}$ cells (Fig. 4a–c).

Before differentiation, the WT and mutant mESCs seemed indistinguishable; however, upon differentiation into cortical neurons, a significant decrease in cell viability was observed in the mutant cells (Fig. 4d). Although increased cell death, affecting mainly non-neural cells, is normally observed during early neuronal differentiation [29], on D3 there was a very significant decrease in the number of $Ex1^{+/-}$ cells (39% relative to WT, $P = 0.0012$) and $Ex10^{+/-}$ cells (14% relative to WT, $P = 6.7 \times 10^{-7}$),

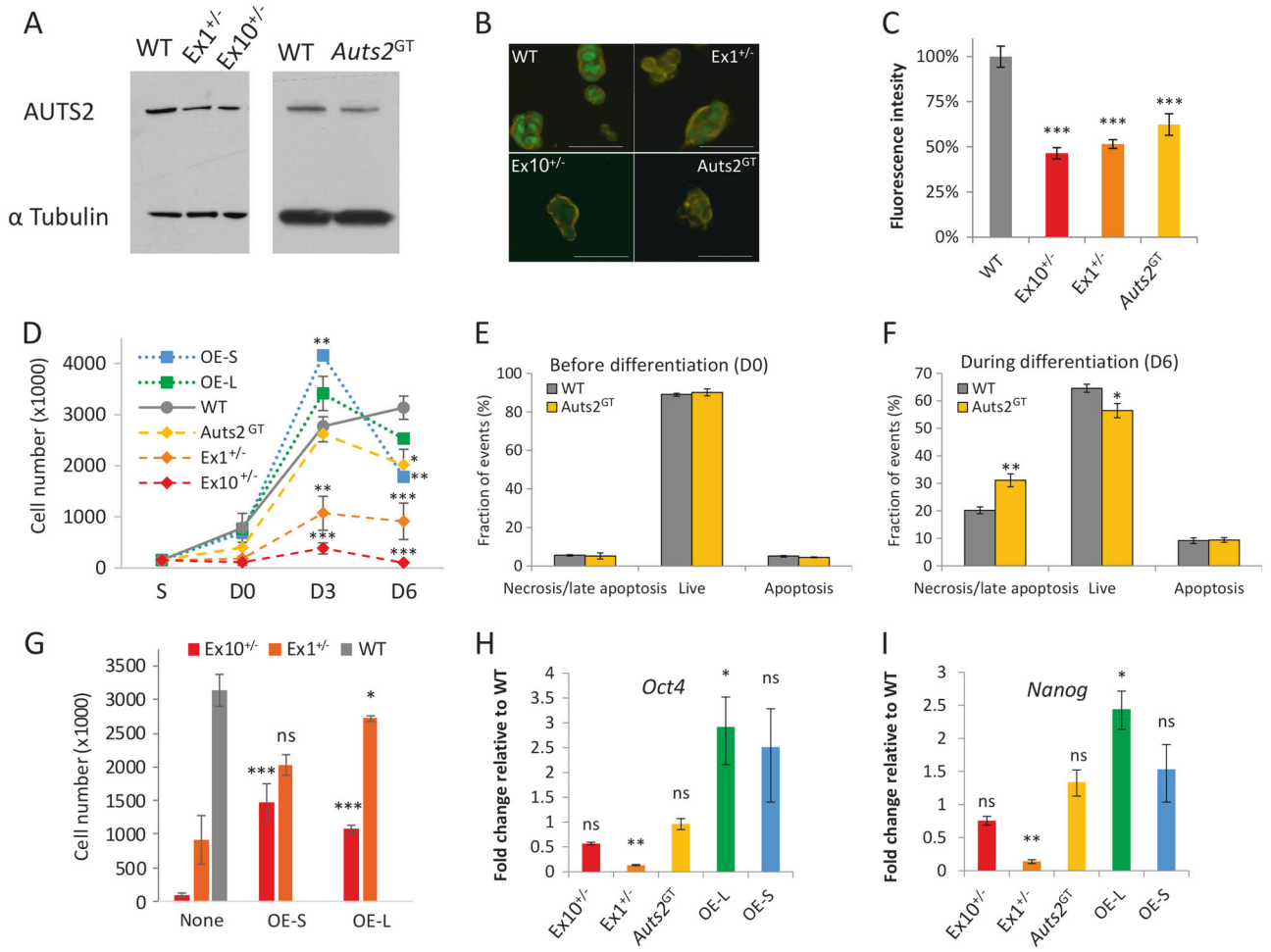

Fig. 4 Mutations in *Auts2* cause increased cell death during in vitro corticogenesis. **a** Western blotting analysis with antibody against the C terminus of AUTS2. α-Tubulin was used as a control for protein loading. **b** Immunostaining of AUTS2 in WT mESCs and cells with heterozygote mutations ($Ex^{+/-}$, $Ex10^{+/-}$, and $Auts2^{GT}$). Green is AUTS2 and yellow is tubulin. Scale bar = 50 µm. **c** Quantification of the AUTS2 signal intensity in mESCs shows a significant reduction (~50%) of AUTS2 in heterozygote mutant cells. Values are average fluorescence intensity (±SEM). **d** The number of viable cells during in vitro corticogenesis in heterozygote mutant cells and cells overexpressing the long or short human *AUTS2* transcripts. Values are the average numbers of live cells (unstained by Trypan Blue) in three independent experiments (±SEM). S, starting number of cells before differentiation ($n = 150{,}000$). **e**, **f** Quantification of Necrosis/late apoptotic, live, and early apoptotic cells before differentiation (D0) and after 6 days of differentiation (D6) of WT and $Auts2^{GT}$ cells. The cells were stained with Annexin V–FITC/propidium iodide and analyzed by flow cytometry. Values are the average of three (D0) or six (D6) independent measurements ± SEM. **g** Overexpression of human *AUTS2* transcripts rescue the reduction in cell viability caused by mutation in *Auts2*. Values are the average numbers of viable cells on D6 in three independent experiments (±SEM). **h**, **i** Pluripotency marker genes in mESCs are influenced by *AUTS2*. The relative expression of Oct4 and *Nanog* in mESCs with heterozygote mutations and cells overexpressing the long or short human *AUTS2* transcripts (ns, not significant; $*P < 0.05$; $**P < 0.01$, $***P < 0.001$; two-sided *t*-test)

which became more severe on D6 (29% and 3%, respectively). $Auts2^{GT}$ cells showed a moderate but significant decrease in cell viability only on D6 (64% relative to WT, $P = 0.017$). As most $Auts2^{GT}$ cells survived (unlike $Ex1^{+/-}$ and $Ex10^{+/-}$), we decided to quantify and determine the mechanism that causes the elevated cell death in $Auts2^{GT}$ cells using the Annexin V and PI double staining followed by fluorescence-activated cell sorting analysis (Fig. 4e, f). Annexin V was used to quantify the number of cells undergoing early apoptosis (PI negative, Annexin V positive) and late apoptosis/necrosis (PI and Annexin V double positive) for cells at D0 and D6. Before differentiation, there was no statistical difference between WT and $Auts2^{GT}$ cells (all $P > 0.05$) (Fig. 4e). However, on D6 we observed a significant decrease in live cells ($P = 0.026$) and a significant increase in the fraction of cells going through a process of late apoptosis/necrosis in $Auts2^{GT}$ cells (31% vs. 20%; $P = 0.0033$) (Fig. 4f). No significant difference was observed for the fraction of early apoptotic cells ($P = 0.89$).

To confirm that *AUTS2* is essential for neuronal differentiation, we used two constructs with the short and long

human *AUTS2* transcripts. We first transfected the constructs to WT mESCs and monitored the cell numbers during differentiation (Fig. 4d). The overexpression of the two transcripts was validated by quantitative PCR (qPCR) (Fig. S5D). Cell numbers were not significantly different between WT and the cells overexpressing the long transcript (OE-L) ($P = 0.11$); however, there was a significant increase in the number of cells overexpressing the short transcript (OE-S) on D3 (150% relative to WT, $P = 0.0018$) and a significant decrease in cell numbers on D6 (57% relative to WT, $P = 0.0050$) (Fig. 4d). We next tested whether overexpression of the two transcripts (OE-L and OE-S) can rescue the decrease in cell numbers observed in D6 for $Ex1^{+/-}$ and $Ex10^{+/-}$. Indeed, a significant increase in cell numbers was observed for $Ex10^{+/-}$ OE-S (14.8-fold, $P = 0.0018$) or the long transcripts (10.8-fold, $P = 0.0018$) and for $Ex1^{+/-}$ OE-L (3.0-fold, $P = 0.0018$) (Fig. 4g). As $Ex1^{+/-}$ and $Ex10^{+/-}$ do not survive the differentiation stages, we hypothesized that mutations in *Auts2* may affect self-renewal or differentiation. We tested whether there are changes in expression of pluripotency genes (*Nanog* and *Oct4*) on D0 by qPCR (Fig. 4h, i). A significant reduction in *Nanog* and *Oct4 was* observed only in $Ex1^{+/-}$ cells (*Nanog*: 14% relative to WT, $P = 0.0036$; *Oct4*: 13% relative to WT, $P = 0.0014$). Correspondingly, a significant increase in the expression of *Nanog* and *Oct4* was observed only in cells overexpressing the long Auts2 transcript (*Nanog*: 240% relative to WT, $P = 0.045$; *Oct4*: 290% relative to WT, $P = 0.047$). Thus, the long transcript, expressed in ES cells before differentiation (Fig. 1c), has the most significant effect on pluripotency.

Gene expression analysis of $Auts2^{GT}$ cells reveals upregulation of polycomb targets involved in neurogenesis

The results of our reporter assay experiments suggest that the first exons of *Auts2*, which are upstream to the gene trap, may functions as a transcription repressor (Fig. 3d). To study the effect of the gene-trap mutation on gene expression, we performed differential expression analysis in three different stages of neuronal differentiation (D0, D6, and D12). We compared between WT and $Auts2^{GT}$ cells, and counted the number of DE genes with at least a 1.5-fold change in expression and a nominal significance ($P < 0.05$). We found that the number of DE genes increased gradually with differentiation (from 24 to 134 and 175 at D0, D6, and D12, respectively). Furthermore, on D0 and D6, there were significantly more upregulated than downregulated genes in $Auts2^{GT}$ (D0: 21 vs. 3, $P = 2.7 \times 10^{-4}$; D6: 87 vs. 47, $P = 6.9 \times 10^{-4}$). However, after 12 days of differentiation, the opposite was observed (52 vs. 123 $P = 8.1 \times 10^{-8}$) (Fig. 5a).

Given that AUTS2 interacts with polycomb proteins, we next tested whether the DE genes are polycomb targets. Many polycomb target genes, such as the *Hox* genes, have been reported to bear both active (H3K4me3) and repressive (H3K27me3) histone marks (bivalent state) in mammalian ESCs [40, 41]. Based on data previously published [42], we calculated the proportion of genes having repressive, active, or bivalent marks in mESCs for each group of genes up- or downregulated on days 6 and 12, as well as for all the genes that were profiled. We found significantly more genes with bivalent marks among the upregulated genes in $Auts2^{GT}$ cells than among the downregulated ones (for D6, $P = 1.2 \times 10^{-4}$, odds ratio [OR] = 6.9, Fisher's exact test; for D12, $P = 1.7 \times 10^{-8}$, OR = 9.9; Fig. 5b).

The genome-wide targets of PRC1 and PRC2 were previously mapped in mESCs by ChIP-Seq [43, 44]. We tested the overlap of genes upregulated in $Auts2^{GT}$ cells with target genes of Ring1b—an essential component of PRC1 that mediates the monoubiquitination of H2A. In addition, we tested the overlap of genes upregulated in $Auts2^{GT}$ cells with genes marked with H2AK119u1 at their transcription start sites, as well as the targets of two PRC2 components Ezh2 and Suz12 (Fig. 5c). To test the significance of the enrichment, we compared the degree of overlap for genes upregulated or downregulated in $Auts2^{GT}$ cells relative to all other genes. We found a significant overlap for genes upregulated at D6 and D12 in $Auts2^{GT}$ cells for all markers (Fig. 5c). At D6, the strongest enrichment was found for genes that are targets of Suz12 (OR = 8.11; $P = 8.9 \times 10^{-17}$), followed by target genes of Ring1b (OR = 8.08; $P = 2.0 \times 10^{-14}$). At D12, the strongest enrichment was found for genes marked by H2AK119u1 (OR = 46.35, $P = 5.4 \times 10^{-14}$) and by Ezh2 (OR = 33.58; $P = 8.9 \times 10^{-17}$).

We next tested the enrichment of biological processes in the DE genes (Fig. 5d). The downregulated genes on D6 were not significantly enriched for any term, but the downregulated genes on D12 were enriched for "muscle contraction" (false discovery rate (FDR) corrected $P = 6.8 \times 10^{-5}$) and "regulation of apoptotic process" (FDR corrected $P = 0.0010$). The upregulated genes on D6 were enriched for terms such as "regulation of multicellular organismal development" (FDR corrected $P = 1.3 \times 10^{-13}$) and "neurogenesis", and on D12 the enrichment was for "synapse organization" (FDR corrected $P = 1.9 \times 10^{-5}$) and for "generation of neurons" (FDR corrected $P = 8.5 \times 10^{-5}$).

AUTS2 affects the timing of neuronal differentiation during in vitro corticogenesis

The upregulation of genes involved in neurogenesis and synapse organization in $Auts2^{GT}$ cells may indicate premature neuronal differentiation. If indeed neuronal

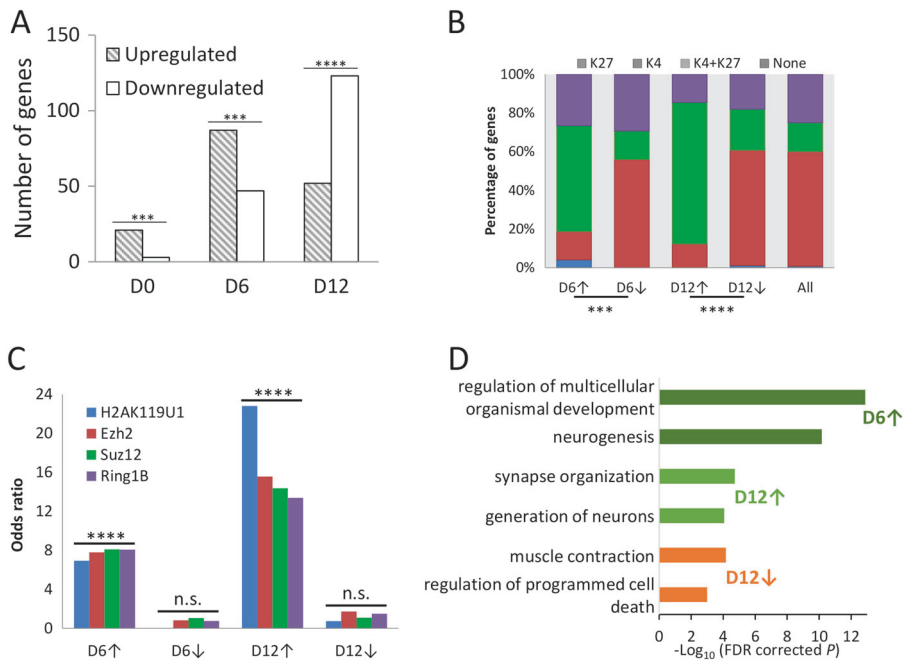

Fig. 5 Gene expression analysis of *Auts2*^{GT} cells during in vitro corticogenesis. **a** The number of genes upregulated or downregulated in *Auts2*^{GT} cells at different stages of differentiation (***$P < 0.001$, ****$P < 0.0001$, exact binomial test). **b** Distribution of genes with active H3K4me3 (K4) mark, repressive H3K27me3 (K27) mark, or bivalent marks containing both active and repressive marks (K4 + K27) in genes upregulated (↑) or downregulated (↓) in day 6 (D6) and 12 (D12). Data are based on histone marks in mESCs from a previously published study [42]. **c** Analysis of the overlap between upregulated and downregulated genes in *Auts2*^{GT} cells and gene targets of PRC1 (*Ring1B*, *H2AK119u1*) and PRC2 (*Suz12*, *Ezh2*). The enrichment is shown by odds ratios that was calculated based on the proportion of overlap with upregulated vs. downregulated genes at both day 6 and 12 (ns, nonsignificant, ***$P < 0.001$, ****$P < 0.0001$, Fisher's exact test). **d** Biological processes enriched in genes upregulated or downregulated in *Auts2*^{GT} cells at different stages of differentiation

differentiation is accelerated in *Auts2*^{GT} cells, gene expression of *Auts2*^{GT} cells might precede that of WT cells during differentiation. To examine this possibility, we compared the difference between *Auts2*^{GT} and WT cells on D6 (*Auts2*^{GT}/WT fold change) to the changes in gene expression between D12 and D6 in WT cells (D12/D6 fold change). The analysis was done for all genes showing a nominal significant difference ($P < 0.05$) in gene expression at D6 between *Auts2*^{GT} and WT cells. Consistent with premature differentiation of *Auts2*^{GT} cells, we observed a very strong linear correlation between the changes in gene expression in the *Auts2*^{GT} cells and changes in expression during differentiation ($R^2 = 0.62$, $P = 1.8 \times 10^{-195}$) (Fig. 6a), indicating that differentiation is accelerated in the *Auts2*^{GT} cells. We next asked whether there are genes that show differential expression, which is not explained simply by the accelerated differentiation and may represent more direct transcriptional targets. To do so, we calculated the residuals (the difference between the measured value and the predicted value of the linear regression model) and identified outlier genes that were more than 1.5 times the interquartile range (Fig. 6b). We identified 57 outlier genes, with 24 genes showing lower expression in *Auts2*^{GT} cells relative to the predicted value and 33 genes with higher expression (Table S4). To test whether the outlier genes are enriched with AUTS2 targets, we compared the list with genes bound by AUTS2 in ChIP-Seq of mouse fetal brain [28]. Out of the 57 genes, 9 were found to be AUTS2 targets, which is a significant enrichment (overlap of 15.8%, when 5.8% is expected by chance, $P = 0.0015$, Hypergeometric test). However, the degree of overlap for the outlier genes (15.8%) was not significantly different than the overlap for non-outlier DE genes (9.0%, OR = 1.75, $P = 0.17$, Fisher's exact test).

The 24 genes with lower-than-expected expression were not significantly enriched for any biological process (FDR corrected $P > 0.05$). The two most extreme outlier genes with lower-than-expected expression in *Auts2*^{GT} cells are *Pou3f2* and *Epha3*. *Pou3f2* (*Brn2*) has a role in production of fate-committed cortical neuronal precursors and is required for upper-layer neural specification and migration [45]. *Epha3* is involved in axon guidance and synaptic development [46, 47]. The 33 genes with higher expression than expected were significantly enriched for "blood vessel development" (FDR corrected $P = 4.7 \times 10^{-4}$) and "mesoderm development" (FDR corrected $P = 6.6 \times 10^{-4}$). The two most extreme outlier genes upregulated in *Auts2*^{GT} cells, *Mest* (*Peg1*) and *T* (*Brachyury*), are mesodermal

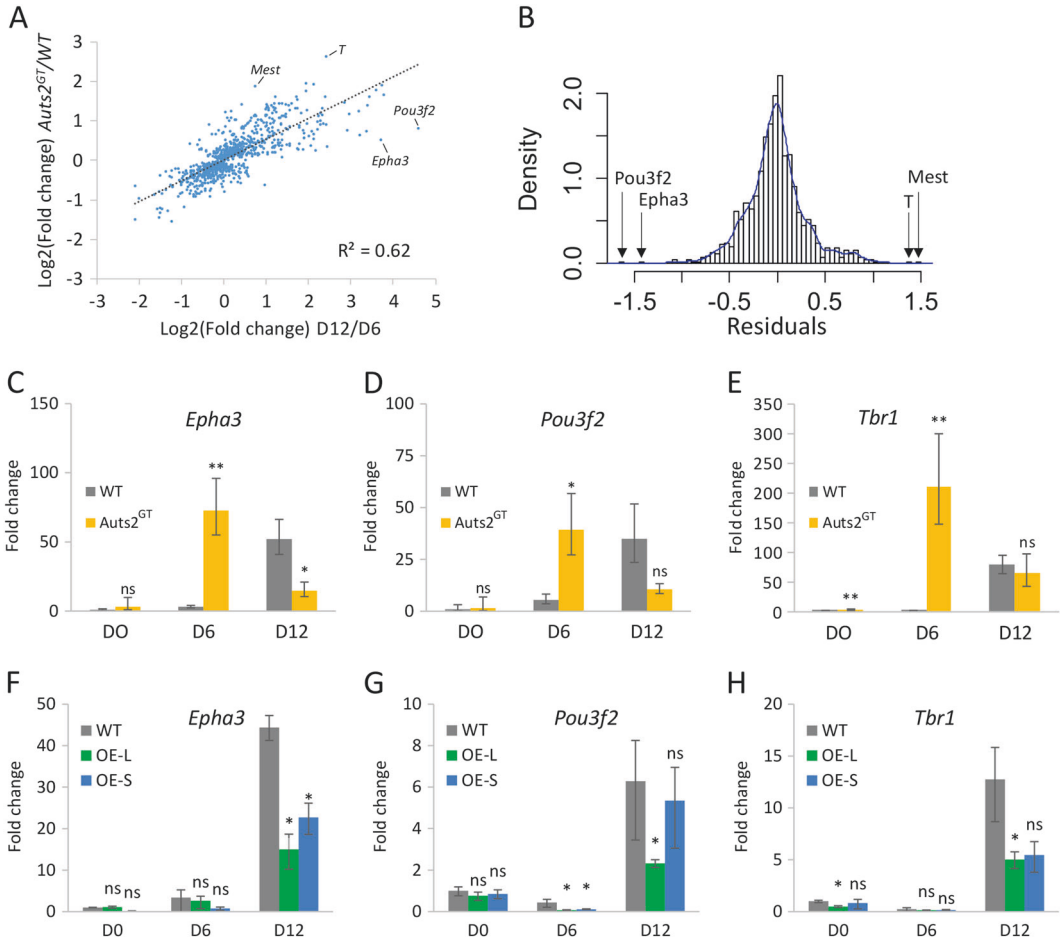

Fig. 6 AUTS2 affect the timing of neuronal differentiation. **a** The differential gene expression between *Auts2*^{GT} and WT cells (D6) is plotted as a function of the change in gene expression between D12 and D6 in WT cells. Values are the average $\log_2$ fold change for all genes showing a nominal significant ($P < 0.05$) difference in gene expression at D6 between *Auts2*^{GT} and WT cells. The dotted line is the linear regression fitted line. R^2 is the proportion of variance explained. **b** Histogram and distribution (blue line) of the residuals from the linear regression line (shown in **a**). The four most extreme outlier genes are indicated by arrows. **c–e** Early upregulation of neuronal markers in *Auts2*^{GT} cells. Relative expression of *Epha3*, *Pou3f2*, and *Tbr1* mRNA in different differentiation stages of WT and *Auts2*^{GT} cells. **f–h** Downregulation of neuronal markers in cell overexpressing *AUTS2*. Relative expression of *Epha3*, *Pou3f2*, and *Tbr1* mRNA in different differentiation stages. Quantitative values are the average of three measurements by quantitative PCR ± SEM (ns, nonsignificant; $*P < 0.05$; $**P < 0.01$; two-sided *t*-test)

differentiation markers [48, 49]. Thus, *Auts2* might be involved in activation of specific cortical neuronal genes and repression of mesodermal genes.

The gene expression analysis suggests that *Auts2*^{GT} cells differentiate faster than WT cells. To validate the role of AUTS2 in differentiation timing, we first replicated the early expression of the two neuronal genes *Pou3f2* and *Epha3* by qPCR. We found that the two genes are highly expressed in *Auts2*^{GT} cells already at D6 ($P = 0.024$, $P = 0.041$, respectively), whereas in WT cells similar level of expression was found only later at D12 (Fig. 6c, d). We then measured the expression of *Tbr1*, which is a marker for postmitotic cortical neurons [50]. Gaspard et al. [29] considered *Tbr1* to be the most valuable marker for efficient dorsalization in their differentiation protocol for cortical neurons from mESCs, showing increased expression after D7 with a maximum reached on D12. In agreement with that study [29], *Tbr1* expression in WT cells was very low at D6 and was upregulated at D12 (Fig. 6e). However, in *Auts2*^{GT} cells, *Tbr1* was upregulated already at D6 ($P = 0.0012$) (Fig. 6e).

We next wanted to test the possibility that the other mutations that we examined (Ex1$^{+/-}$ and Ex10$^{+/-}$) also result in accelerated neuronal differentiation. This is supported by the lower expression of the pluripotency markers, *Nanog* and *Oct4*, in Ex1$^{+/-}$ cells (Fig. 4h, i). However, as Ex1$^{+/-}$ and Ex10$^{+/-}$ cells do not survive the differentiation process (Fig. 4d), we used cells at D0 and D3, and measured the expression of *Pou3f2* and *Epha3*. Consistent with accelerated differentiation, there was a higher expression of *Epha3* at D3, which was significant only for Ex10$^{+/-}$ cells relative to WT cells ($P = 0.033$) (Fig. S6A-B).

Finally, we tested whether overexpression of *AUTS2* transcripts will have an opposite effect by delaying neuronal differentiation. Consistent with the role of AUTS2 in controlling neuronal differentiation, we found on D12 an increase of Nestin staining in cells overexpressing the long and short *AUTS2* transcripts relative to WT cells (Figure S6C). Similarly, cells overexpressing the long *AUTS2* transcript had a reduced expression of *Epha3*, *Pou3f2*, and *Tbr1* on D12 relative to WT cells ($P = 0.019$, 0.039, and 0.043, respectively) (Fig. 6f–h). In cells overexpressing the short isoform, there was only a significant reduction in expression of the *Epha3* gene on D12 ($P = 0.02$) (Fig. 6f–h). This suggests that overexpression of *AUTS2*, particularly the long transcript, inhibits neuronal differentiation.

Discussion

Our study demonstrates that the precise expression of AUTS2 isoforms is critical for the timing of neuronal differentiation. By focusing on the role of *AUTS2* during neuronal differentiation from mESCs using WT and genetically manipulated mESCs, we were able to make three major observations.

First, we demonstrate that there is an isoform switch during neuronal differentiation. The long AUTS2 isoform, expressed in undifferentiated cells, is replaced by a short isoform during neuronal differentiation. The short isoform is controlled by an alternative promoter, which is conserved between human and mouse. We also identified alternative exon skipping events that are specific to the developing brain in both mouse and human. This is reminiscent of the SET protein, which shows a rapid isoform shift during early neuronal differentiation through alternative promoters [51].

Second, we find that the effect of AUTS2 on transcription is highly influenced by the isoform and the cellular context. The results suggest that the C terminus of AUTS2, which is included in the short isoform, is responsible for transcription activation, whereas the N-terminal region, which is specific to the long isoform, is able to repress transcription. We assume that the activity of AUTS2 isoforms depends on co-factors that are differently expressed between cells and on factors that differentially interact with the two isoforms. One such factor is PCGF3, which is expressed in N2A cells but not in HEK293, and, based on our Y2H screen, interacts only with the long AUTS2 isoform. Indeed, we found that overexpression of PCGF3 in HEK293 cells shifted the function of the long AUTS2 isoform from activation to repression and knockdown of PCGF3 in N2A cells resulted in loss of transcription repression. Our results are also in agreement with a previous study that demonstrated that the N terminus of AUTS2 is unable to recruit the co-activator P300 [26]. The two isoforms interact with SF3B1, suggesting a role of AUTS2 in chromatin-splicing interactions [39]. Consistent with the role of AUTS2 in chromatin-splicing interactions, we found that the genes co-expressed with *AUTS2* in the developing brain are enriched for chromatin modification, RNA processing, and RNA splicing.

Third, the manipulation of AUTS2 levels by different types of heterozygote mutations and overexpression of the two isoforms revealed that AUTS2 is critical for the regulation of timing and speed of neuronal differentiation. The most significant effect was for the long transcript. Overexpression of the long transcript led to increased expression of pluripotency markers and delayed neuronal differentiation. Mutation by a gene trap that leads to a transcript terminated prematurely, but that contains the region specific to the long transcript, resulted in accelerated differentiation.

The timing of neurogenesis is crucial for determining the size of the initial stem cell pool and the size of the brain. Thus, our results suggest that the microcephaly seen in individuals with *AUTS2* mutations may result from accelerated differentiation to neurons that produces deficiencies in progenitor cells at later stages of neurogenesis. In this context, it is interesting to note that similar to other genes that control cerebral cortex size, *AUTS2* was subject to selection during the evolution of modern humans [52]. Our gene expression analysis highlights two neuronal genes that may be direct targets of AUTS2: *Epha3* and *Pou3f2*. A previous ChIP-Seq study in mouse fetal brain showed that AUTS2 binds the gene region of *Epha3* [28]. *Pou3f2* is an interesting potential target of AUTS2, as it is a transcription factor that plays a key role in neuronal differentiation that was found to be one of three factors that is sufficient to convert mouse embryonic and postnatal fibroblasts into functional neurons in vitro [53].

Taken together, our results indicate that AUTS2 is a key gene regulating the proper timing of neuronal differentiation in vitro. The complex functions of AUTS2 isoforms offer new mechanistic insights into NDDs, and how loss-of-function mutations in different parts of a gene can cause different symptoms related to the function of different isoforms.

Acknowledgements This research was supported by German-Israeli Foundation for Scientific Research and Development (GIF) (grant number 2224/2009 to SS), by the Israel Science Foundation (grant numbers 688/12 to SS, 575/17 to SS and 1140/17 to EM), and supported in part by 'Investissements d'Avenir' program (Labex Biopsy, ANR-11-IDEX-0004-02 to MG) and Fondation Lejeune (to MG).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485:237–41.
2. Hamdan FF, Srour M, Capo-Chichi J-M, Daoud H, Nassif C, Patry L, et al. De novo mutations in moderate or severe intellectual disability. PLoS Genet. 2014;10:e1004772.
3. Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. Mol Psychiatry. 2013;18:1054–6.
4. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet. 2014;94:677–94.
5. Ronan JL, Wu W, Crabtree GR. From neural development to cognition: unexpected roles for chromatin. Nat Rev Genet. 2013;14:347–59.
6. Suliman R, Ben-David E, Shifman S. Chromatin regulators, phenotypic robustness, and autism risk. Front Genet. 2014;5:81.
7. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515:209–15.
8. Ernst C. Proliferation and differentiation deficits are a major convergence point for neurodevelopmental disorders. Trends Neurosci. 2016;39:290–9.
9. Casanova EL, Casanova MF. Genetics studies indicate that neural induction and early neuronal maturation are disturbed in autism. Front Cell Neurosci. 2014;8:397.
10. Kaushik G, Zarbalis KS. Prenatal neurogenesis in autism spectrum disorders. *Front Chem*. 2016;4:12. https://doi.org/10.3389/fchem.2016.00012.
11. Packer A. Neocortical neurogenesis and the etiology of autism spectrum disorder. Neurosci Biobehav Rev. 2016;64:185–95.
12. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014;506:179–84.
13. Shohat S, Ben-David E, Shifman S. Varying intolerance of gene pathways to mutational classes explain genetic convergence across neuropsychiatric disorders. Cell Rep. 2017;18:2217–27.
14. Zhu X, Need AC, Petrovski S, Goldstein DB. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. Nat Neurosci. 2014;17:773–81.
15. Sultana R, Yu C-E, Yu J, Munson J, Chen D, Hua W, et al. Identification of a novel gene on chromosome 7q11. 2 interrupted by a translocation breakpoint in a pair of autistic twins. Genomics. 2002;80:129–34.
16. Kalscheuer VM, FitzPatrick D, Tommerup N, Bugge M, Niebuhr E, Neumann LM, et al. Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation. Hum Genet. 2007;121:501–9.
17. Huang X-L, Zou YS, Maher TA, Newton S, Milunsky JM. A de novo balanced translocation breakpoint truncating the autism susceptibility candidate 2 (AUTS2) gene in a patient with autism. Am J Med Genet Part A. 2010;152:2112–4.
18. Ben-David E, Granot-Hershkovitz E, Monderer-Rothkoff G, Lerer E, Levi S, Yaari M, et al. Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. Hum Mol Genet. 2011;20:3632–41.
19. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry. 2009;15:637–46.
20. Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. 2010;6:e1000962.
21. Amarillo IE, Li WL, Li X, Vilain E, Kantarci S. De novo single exon deletion of AUTS2 in a patient with speech and language disorder: a review of disrupted AUTS2 and further evidence for its role in neurodevelopmental disorders. Am J Med Genet A. 2014;164A:958–65.
22. Beunders G, Voorhoeve E, Golzio C, Pardo LM, Rosenfeld Ja, Talkowski ME, et al. Exonic deletions in AUTS2 cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. Am J Hum Genet. 2013;92:210–20.
23. Bedogni F, Hodge RD, Nelson BR, Frederick EA, Shiba N, Daza RA, et al. Autism susceptibility candidate 2(Auts2) encodes a nuclear protein expressed in developing brain regions implicated in autism neuropathology. Gene Expr Patterns. 2010;10:9–15.
24. Oksenberg N, Stevison L, Wall JD, Ahituv N. Function and regulation of AUTS2, a gene implicated in autism and human evolution. PLoS Genet. 2013;9:e1003221.
25. Hori K, Nagai T, Shan W, Sakamoto A, Taya S, Hashimoto R, et al. Cytoskeletal regulation by AUTS2 in neuronal migration and neuritogenesis. Cell Rep. 2014;9:2166–79.
26. Gao Z, Lee P, Stafford JM, von Schimmelmann M, Schaefer A, Reinberg D. An AUTS2–Polycomb complex activates gene expression in the CNS. Nature. 2014;516:349–54.
27. Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell. 2012;45:344–56.
28. Oksenberg N, Haliburton GDE, Eckalbar WL, Oren I, Nishizaki S, Murphy K, et al. Genome-wide distribution of Auts2 binding localizes with active neurodevelopmental genes. Transl Psychiatry. 2014;4:e431.
29. Gaspard N, Bouschet T, Herpoel A, Naeije G, van den Ameele J, Vanderhaeghen P. Generation of cortical neurons from mouse embryonic stem cells. Nat Protoc. 2009;4:1454–63.
30. Gaspard N, Bouschet T, Hourez R, Dimidschstein J, Naeije G, van den Ameele J, et al. An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature. 2008;455:351–7.
31. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8:2281–308.
32. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9:215–6.
33. Ge H, Liu Z, Church GM, Vidal M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet. 2001;29:482–6.
34. Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. Genome Res. 2002;12:37–46.
35. Deane CM, Salwiński Ł, Xenarios I, Eisenberg D. Protein interactions: two methods for assessment of the reliability of high throughput observations. Mol Cell Proteomics. 2002;1:349–56.
36. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science. 2003;302:449–53.
37. Chen J, Xu H, Aronow BJ, Jegga AG. Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics. 2007;8:392.
38. Isono K, Mizutani-Koseki Y, Komori T, Schmidt-Zachmann MS, Koseki H. Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. Genes Dev. 2005;19:536–41.

39. Kfir N, Lev-Maor G, Glaich O, Alajem A, Datta A, Sze SK, et al. SF3B1 association with chromatin determines splicing outcomes. Cell Rep. 2015;11:618–29.
40. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125:315–26.
41. Harikumar A, Meshorer E Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO Rep*. 2015;16:1609–19.
42. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.
43. Endoh M, Endo TA, Endoh T, Isono K, Sharif J, Ohara O, et al. Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genet. 2012;8:e1002774.
44. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 2008;4:e1000242.
45. McEvilly RJ, de Diaz MO, Schonemann MD, Hooshmand F, Rosenfeld MG. Transcriptional regulation of cortical neuron migration by POU domain factors. Science. 2002;295:1528–32.
46. Brennaman LH, Zhang X, Guan H, Triplett JW, Brown A, Demyanenko GP, et al. Polysialylated NCAM and ephrinA/EphA regulate synaptic development of GABAergic interneurons in prefrontal cortex. Cereb Cortex. 2013;23:162–77.
47. Nishikimi M, Oishi K, Tabata H, Torii K, Nakajima K. Segregation and pathfinding of callosal axons through EphA3 signaling. J Neurosci. 2011;31:16251–60.
48. Sado T, Nakajima N, Tada M, Takagi N. A novel mesoderm-specific cDNA isolated from a mouse embryonal carcinoma cell line (embryonal carcinoma cell/cDNA/in situ hybridization/mesoderm/mouse embryo). Dev Growth Differ. 1993;35:551–60.
49. Martin BL, Kimelman D. Brachyury establishes the embryonic mesodermal progenitor niche. Genes Dev. 2010;24:2778–83.
50. Bedogni F, Hodge RD, Elsen GE, Nelson BR, Daza RAM, Beyer RP, et al. Tbr1 regulates regional and laminar identity of post-mitotic neurons in developing neocortex. Proc Natl Acad Sci USA. 2010;107:13129–34.
51. Edupuganti RR, Harikumar A, Aaronson Y, Biran A, Sailaja BS, Nissim-Rafinia M, et al. Alternative SET/TAFI promoters regulate embryonic stem cell differentiation. Stem Cell Reports. 2017;9:1291–303.
52. Oksenberg N, Ahituv N. The role of AUTS2 in neurodevelopment and human evolution. Trends Genet. 2013;29:600–8.
53. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature. 2010;463:1035–41.